

**Response to Intervention  
Cohort training session:  
Process Evaluation: Is it Working?**

**Session 2**

**July 12 - July 13, 2012  
Virginia Beach, Virginia  
Sponsored by the  
Virginia Department of Education  
Office of Student Services**

Response to Intervention  
Cohort training session: Process Evaluation: Is it Working?

VA\_DOE\_RTI\_Session\_1\_7\_13\_2012\_PART2.mp3

Validity and Accountability: Planning Ahead      –      10:30 a.m. – Noon

Dr. Kenneth Hinson, Supervision of Program Evaluation  
Prince William County Public Schools

Dr. Paul Parker, Supervisor of Testing  
Prince William County Public Schools

Dr. Nancy Cary, Coordinator of Program Evaluation  
Prince William County Public Schools

00:00:00

Dr. Cave: Ready? Everybody ready? Okay good. All right, so the next oh a few more people coming in, the next hour and a half we sort of designated Prince William County is so large that they have their own huge office of program evaluation and accountability. So basically the question comes up for whatever they do in Prince William, these folks will say is it working or is it not? And the people that have come today, Ken and Paul, they're like the dream team of accountability. It's amazing and all of the teachers whenever you hear them, or the Title 1 people they'll always say well let's ask Ken if that's gonna work, or let's talk to Paul and see if we can figure this out. They always ask those really good guiding questions of things like, are you know, how do I know that what you're showing me really is showing me what you think it's going to be showing me, or things like that. So it always stretches their neurons to have to work with them, but we're always really appreciative of being able to work with them.

00:01:00

Dr. Cave: So basically what we're gonna talk about is, is it working? Program and process evaluation is it working? And it's going to be referring to the VTSS benchmarks, so Prince William thinking ahead has said once they got the benchmarks, okay if this is what we're supposed to implement how do we know that we're really implementing it the way it's intended to be implemented. So without further ado I'll turn it over this is Ken this is Paul.

Dr. Hinson: Good morning. Again I'm Ken Hinson I run the program evaluation shop in Prince William County Public Schools. Been there now for about five years. Do the same kind of job in Fairfax for five years too. Benchmarks there are a lot of those aren't there. I mean a bunch of them, but Paul and I, we had a third member but unfortunately

## Response to Intervention

### Cohort training session: Process Evaluation: Is it Working?

she had family issues to take care of so she won't be with us today, Nancy Cary. But this is what we're gonna cover today, talk a little bit about process evaluation.

00:02:01

Dr. Hinson: And of course tiered system of supports benchmarks, ways to evaluate the processes involved in attaining and knowing that you have met, or that you are trying to meet the benchmarks. Some comments at the later time, Paul will talk about the aspects of validity as it relates to assessment, some following comments, and once the both of us are done this morning we have an activity that will take I don't know, 15, 20, 30 minutes depending on the level of interest and participation. Now each of you should have at your tables at least a version of this PowerPoint. Note pages if you will, you don't have one?

00:03:00

Dr. Hinson: Right off the bat evaluation takes of different definitions, connotations, especially in the educational arena. As a program evaluator we basically look at evaluation as involving systematic inquiry that basically is done to help us either improve programs, interventions, personnel action, and all actions that people do leading up to working these particular areas out. Process evaluation should be looked upon as putting emphasis on the early stages of the implementation process, and that could be defining goals and objectives, assigning roles and responsibilities to all stakeholders involved.

00:04:04

Dr. Hinson: We're looking at conducting training as necessary, professional development, monitoring the efforts. That's a big word, big phrase, depending upon where you are and the resources you have at hand that will take on different levels of action. And basically collecting feedback on how well the stakeholders, how well things are going based upon what stakeholders give you in the way of feedback, and what we mean there is stakeholders who are managing and running the program, stakeholders who are benefactors of the program, and any other stakeholders who have a vested interest with respect to the program or the intervention. Again we're gonna talk about the tiered system of supports, we talked about there being a lot of them, and the process I'm gonna go through you're gonna find that it's gonna seem repetitive as far as the format goes. But hopefully the content will provide some information, bits of ways to do things that will help you at whatever level you are in your particular divisions.

00:05:08

Dr. Hinson: The benchmarks are listed in the big table you have, what I have done with those particular benchmarks. I have eight of them that I will talk about today. I have

## Response to Intervention

### Cohort training session: Process Evaluation: Is it Working?

reframed them in the form of a question for evaluation purposes, and the first question comes from the leadership area, the emerging stage of that particular benchmark. To what extent has a clearly written vision statement been created and shared at the division and school level? This is gonna be sort of guinea pig slot, I won't go in to very much detail there, but to get to answering that first question there probably are several other questions that should come to mind in order to be able to provide the data and information to lead you to be able to know whether that particular benchmark is being met at some level.

00:06:02

Dr. Hinson: Was development of the vision statement a shared responsibility among stakeholders? Lot of stakeholders would be involved probably. Central office people, school based administrators, teachers, student, parents, community members at large, and once you know who those stakeholders are, a question is this. How did they participate in the process? You know, did you survey them, did you have a focus group, were they all assigned to an advisory group to come up with what took place. And of course the way that you could somewhat verify as to the level or participation, or if they participated, and you'll see this more so when I get into the other slides, is by checking with them through surveys, focus groups, interviews to see if in fact they participated, how they participated, and what part they played overall in getting this particular benchmark worked up as far as the vision statement is concerned.

00:07:12

Dr. Hinson: The next one, to what extent is professional development that supports TSS divided across administrative, operation, instruction, and support levels? Questions you could consider with respect to the larger question, are professional development plans available across a spectrum of stakeholders? What is the availability and what kind of professional development do we have? Is the professional development set up such that you can address unique requirements with respect to individual stakeholders? And an open ended question that can be posed is, is the professional development content relevant to stake holders needs?

00:08:03

Dr. Hinson: And moving on, after receiving professional development is it used to inform/adjust classroom practices or practices of the various administrative offices? Lastly are there any changes to be made with respect to TSS? And now how we're going to collect data to know that we have answers to these key questions, we can obvious review professional development plans, look at attendance logs, we can observe practices in meetings, and of course we can gain stake holder perceptions through focus groups, interviews, and surveys.

## Response to Intervention

### Cohort training session: Process Evaluation: Is it Working?

00:09:03

Dr. Hinson: Packs of questions that you could pose to these stake holders, this would be a survey item, professional development is relevant to my professional needs, indicating whether you strongly agree, agree, disagree, or strongly disagree. Knowledge, skills gained from professional development are used, again a different way of looking at it, all of the time, most of the time, some of the time, none of the time to adjust instructional practices. Various ways you can pose questions in either of these venues, and lastly how can professional development be improved for all involved? Next question, to what extent does the division conduct regular needs assessment and student and staff surveys to assess climate toward improved academic performance of students?

00:10:02

Dr. Hinson: And this is the organizational culture area, the emerging stage of that particular benchmark. Big question, what stakeholders would you survey to collect this information? Again you have the list there of just about everybody involved in the school, and of course the parents themselves. What areas are you going to cover on the survey? And indication of some areas that you could consider, quality of instructional programs, services, and otherwise. The environment in which teaching and learning takes place, how much families are involved in the decision making process, even just participating in school activities, teacher quality is a big area of concern to all. So now how do you collect information, with respect to teachers and this is strictly, strictly an example of the kinds of questions that could be posed, again in the venues of focus groups, interviews, or by way of surveys.

00:11:09

Dr. Hinson: Students, questions posed to students could be as follows: adults at my school care about me, and the student would be asked to strongly agree, agree, disagree, or strongly disagree with that particular statement. Other statements that follow: I feel safe from other students at school, kind of getting into the bullying arena. My teachers expect that all students can do their best, high expectations across the board for all students. I feel that I am getting a good education at my school. Kids could again strongly agree, agree, or disagree with these particular statements. Of course pose a question to these kids basically saying, what do you think could be done to improve your school at this particular time?

00:12:01

Dr. Hinson: So again as I say you follow the format is basically the same, but the information is a little bit different. And I apologize for that, but I think it will sort of reiterate and push the point home as to how you can go about this process. Now how

## Response to Intervention

### Cohort training session: Process Evaluation: Is it Working?

do we collect the data with respect to teachers? My school offers a wide variety of opportunities for parent/community involvement and decision making, what level of agreement do they have with that? I am satisfied with the instructional programs at my school. I have a wide variety of delivery options of getting and using staff development. I as the teacher am satisfied with my teaching and learning environment, and again to the teacher what improvements do you feel should be done or made at your particular school?

00:13:02

Dr. Hinson: Now all the questions that you see that I post thus far and somewhat following are questions that are actually part of the overall Prince William County division wide survey for parents, teachers, and students. It's a smattering of those kinds of questions. With respect to parents, following kinds of questions. I have a good relationship or relationships with staff at my child's school, or even the teacher. Overall I am satisfied with my child's school environment. Teacher quality, my child has good teachers. Overall my child is getting a good education. There are my different ways that I as a parent am able to participate, involve myself with what goes on at my child's school. I feel welcome at the school.

00:14:00

Dr. Hinson: Just a few of the kinds of questions we pose to parents with respect to the climate of the school, the division. Yes sir. We, we survey in the division no less than 4<sup>th</sup> grade, 4<sup>th</sup>, 5<sup>th</sup> graders up through 12<sup>th</sup> grade. Anything that's one through three because of the kinds of questions that are posed, okay. And in what way could your child's school be improved? Next benchmark item, to what extent does the division use a formal system of monitoring student progress that includes universal screening, formative assessment, and other diagnostic processes? Part of the monitoring student progress benchmark. A question you could ask to get to that larger question, what is the formal system of monitoring student progress within our division, within our own schools?

00:15:06

Dr. Hinson: There is a myriad of them, most of you are familiar with what I have listed on the screen right now. Curriculum based measurement, of course we're probably all familiar with AIMSweb. Dynamic indicators of basic early literacy, mClass Math, STAR math, and other formative assessments that may be occurring locally and any combination of various tools and assessment at hand. Now given all that you have to work with, how do you know that the formal assessment system is being used? How do you know if the assessment system is being using within your particular division? Again reviewing reports, sitting in meetings, observing practices, again getting stakeholder

## Response to Intervention

### Cohort training session: Process Evaluation: Is it Working?

input. Are tools for monitoring progress adequate for the task at hand? Are there too many of them?

00:16:02

Dr. Hinson: Depending on where you some people say yes, some say no, some say there are not enough. Are too standardized across the division, depending on how large you are, you know, is one school using one thing, the other school using another, and granted given the context of students sometimes that is necessary, so saying it's standardized is not necessarily a hard and fast rule because know that there are exceptions all the way across. What seems to be working with respect to these tools? What's not working? Do you have a process for adding, subtracting from your inventory? And again how can you improve the overall process with respect to what you're doing in monitoring the student's progress in your division or in your schools? Here is a good one, to what extent are all decisions made using division wide problem solving process, which is consistently implemented with fidelity.

00:17:02

Dr. Hinson: And we all know what fidelity is with respect to marriage right? Or infidelity, but we know that when we talk about fidelity it just means that whatever you're doing is done in accordance with established rules, regulations, processes, etc. In Prince William County most of the time we use PDSA on a general level, plan, do, study, act. Planning is where you develop objective questions and try to predict what you would like to see as outcomes. The doing is the carrying out the plan where you document issues that you have with respect to what's going on. You begin to analyze data with respect to those issues and problems. The study phase you continue analyzing the data, document whatever issues you have, and compare those, what you, what you have with respect to results with the overall plans and objectives that you started out with in the beginning to see if there are any variations or differences. Act, you determine what changes should be made based on all these analyses, and then the process starts all over again.

00:18:05

Dr. Hinson: If you find issues you begin to start the cycle all over again. Something that was alluded to earlier, but a problem solving process related to Rtl, define the problem, analyze the problem, develop the plan, implement the plan, evaluate the plan. Some of you are familiar with that, planning the problem. Determine what makes up the problem and in the analysis phase there you really look to figure out what caused the problem with the data that you can collect. You develop a plan to try and address the problem, solve the issues, implement the plan, put into motion, and at the same time you're putting it into motion you want to monitor how you're progressing with respect to that effort. And you evaluate the plan, in other words you'll see how effective, or how well it

## Response to Intervention Cohort training session: Process Evaluation: Is it Working?

worked. And based upon those results again you probably start that cycle all over again with respect to this particular way of problem solving. And most of you are probably familiar with the [Baldrich ?] model of continuous improvement.

00:19:04

Dr. Hinson: We use that in most of our, I guess elementary schools in Prince William County right now at some level. So how do we know that the processes working are being used in your division or schools, going back through the same format you know. Review documentation, observe practices, what goes on in meetings, quite like the short snippets we saw earlier today. We observed after the fact but that's the sort of thing we want to be looking at. Gaining stakeholders perceptions on what's going on, and in this particular case we can look at these kinds of questions for stakeholders to answer. A division wide problem solving process is used to make decisions, and you can ask the question one way, and that one way is that you can say is it happening?

00:20:03

Dr. Hinson: Where you get a yes or no answer, and if you get a yes answer you could say well we're doing fine, but you may not be doing so well. So to kind of get at the heart of a little bit more extension, or extent, pose the question say well is it all the time, most of the time, some of the time, none of the time. Hopefully none of the time is every there but you get the general idea. Another way to phrase that division wide problem solving process is implemented with fidelity, I strongly agree with that, disagree and whatever level in between. Quality decisions are made using the problem solving process in our area of operation, true statement, I strongly agree all the way down to disagreeing with that particular statement. And always as I've said before you can pose an open ended question to basically see how well your thing is going, or if people feel that there should be improvements made along the way.

00:21:07

Dr. Hinson: To what extent are curriculum intervention resource decisions being made, are made using progress monitoring data, and are they used with fidelity? That goes back to I believe your movies today. The larger question you could ask is, how should curriculum resource decisions be made? How should they be made? Do we involve central office? Do we involve teacher leaders, individual teachers, professional learning communities, grade level teams, content area teams, any combination of these stakeholders you could consider in that effort.

00:22:01

Dr. Hinson: Again how do you know that it's being done, observing, reviewing documentation and data, and collecting information from the stakeholders based upon

## Response to Intervention

### Cohort training session: Process Evaluation: Is it Working?

these kinds of questions. And again these kinds of questions that I'm posing here are not necessarily right or wrong, and not necessarily what you should follow, but it just gives you an idea of the kind of things that you can think about with respect to your own context. Curriculum resource decisions are made using progress monitoring data. Again how much, or how often, to what extent? Decisions are being made are used with fidelity, I agree or I disagree at some level, and ask the stake holders to please provide some information with respect to the kinds of changes they think should be made to the overall process that we have going here. To what extent are evidence based instructional processes that are an appropriate match for the needs of the student used with fidelity?

00:23:04

Dr. Hinson: This is under instruction, sustaining stage, and again you'll see I'm only talking about eight of those benchmarks and there are lots of them, lots of them. To, to kind of start the, the process I would say you need to identify what evidence practices are available and appropriate to meet the student's need. And I just divert here for a moment, we often hear the term evidence based and research based. Evidence based is always research based, but research based is not necessarily evidence based. A researcher can conduct a study on one group, one population, come up with great results, great findings, and may say that that this is what should happen under certain circumstances. That's not evidence based, that's research based, but it's not evidence based.

00:24:01

Dr. Hinson: Evidence based would be a situation where we go across the spectrum, where we research done on a robust situation where researchers over time have shown that we get the same kinds of findings with respect to similar circumstances. That'll let you know it will probably work in your environment. With respect to finding evidence based practices, or tools if you will, the US Department of Education has what's called as the what works clearinghouse. That is at least one place you can go to see if what it is you're using that is purposed to be evidence based you can confirm it there, there are probably other locations. So how do you know the practices are evidence based and appropriate? Again by reviewing documents, checking with those sites that I indicated before, collecting stakeholder information, with the kinds of questions I would pose either on a survey, in a focus group, or in an interview venue.

00:25:08

Dr. Hinson: Here is a good one, to what extent do parents view themselves as partners and advocates in the school division's tiered system of supports, this is under the family and community partnership benchmark at the sustaining stage. Consider these questions in light of trying to answer the big question, how have parents been informed

## Response to Intervention

### Cohort training session: Process Evaluation: Is it Working?

of the division's TSS? Have parents been invited to play active roles in the TSS? How does the division school, or school view parents as partners or advocates, or as advocates, how do parents themselves view what they do with respect to being advocates and partners?

00:26:00

Dr. Hinson: If we expect that they are how do we know? Same litany of steps we followed before, but coming down to the kinds of questions we could pose to parents we can say mark all that apply. You have been made aware of the division's TSS by those various ways or venues, school board meetings, newsletters from the teachers, emails, phone calls, divisions and school websites, back to school nights, parent teacher meetings, so on and so forth. You could ask a parent to answer these particular questions, I was invited to participate in the division's TSS by those various means or others in addition to those. The role that I fulfill for the division's TSS are, you know, think about it, what do you want parents to do? What do the parents want to do themselves? Classroom volunteers, serve as tutors, serve as support team members, advisory team participants.

00:27:03

Dr. Hinson: Or just serve as their child's advocates, I'm just going to check on how well she or he is doing, contact the teacher, contact the principal as is necessary. To what extent to division data demonstrate that the TSS model has led to improved outcomes for students according to multiple measures? This is part of the evaluation plan benchmark. And often times we say that when, when something has led to an outcome in research terms that normally means that it, it's caused it if you will, or had an impact on it happening. And in many cases that's difficult to show if we haven't gone through a very rigorous robust process of research and analysis, but what we normally do is we analyze the data for indicators or trends of improvement for kids.

00:28:00

Dr. Hinson: Normally what we've done in divisions in the state from time to time, we have looked at cohorts, the current results for cohorts and compared them to previous cohort results either one, two, three, four years back however the case may be. We could look at annual results from summative assessments against SOLs, against SAT scores and so forth. And with the state's new growth model what we can do now, we can basically look at a child's performance on the current administration of the standardized test, and compare that with the previous administration and get some sense of growth for that individual student, or as divisions as a whole. Final comments, process evaluation is strictly pretty much tied to the early stages of implementation and that's gonna vary depending upon how large or small you are as a division, and I would

## Response to Intervention Cohort training session: Process Evaluation: Is it Working?

say depending upon all you have to look at with respect to what's needed to be met is choose what's most important and tie that to what your resources will allow you to do.

00:29:11

Dr. Hinson: And lastly process evaluation is tied to the beginning stages, but actually to make sure that your intervention, that your program is working throughout it's lifespan you need to conduct these kinds of assessments, reviews etc. throughout the lifespan of the program. Yes ma'am. That's gonna vary depending on the level, in at my level in Prince William County if we're talking about summative data, you know, SOL results we have a data group, a data team. All the results come through and we put...

00:30:03

Dr. Hinson: Okay. And shared, and you're saying who does it. Right. Yes my office does that and we provide the reports and recommendations, and we share that with of course the, the main stake holders, but it does from division directors, to school principals, to associate superintendents, up to the executive group.

00:31:07

Dr. Hinson: And everybody understands what's there, and we talk about the recommendations, and, and what we normally do in situations like that, if recommendations are tied to a particular organizations we will try to ask them to, this is what we are recommending based up on the findings, do you agree or disagree? Whatever they agree or disagree with we ask them to give us in addition to the report itself a response to the findings of the recommendations. How do we go about what we've said should take place, and all this goes up to the superintendent at the superintendent level, and the executives dealing with that. And that's basically how it's shared with everyone.

00:32:00

Dr. Hinson: Yeah. Yeah, and many of our schools we have parent liaisons, and a role they play is being in contact with parents, and many principals in schools that's their avenue of being in direct contact with parents with, with major issues, major programs of whatever, and how we want to deal with it. Again I talked about the, the back to school nights when I said school board meetings, these kinds of things are discussed at our school board meetings quite often. And we have citizens there, and it's televised so we, that's one venue if you will where parents can become educated with respect to that process. Yes. No, not yet.

00:33:00

## Response to Intervention Cohort training session: Process Evaluation: Is it Working?

Dr. Hinson: We have not been collecting data, okay let me go back and rephrase the overall gist of my discussion. I am a program evaluator, and whatever the issue is I have in my bag of tricks the ability to determine how an approach should be laid out to address the issues. Now we have RtI, we are looking at the tiered system right now, and we have a sub-group if you will that's been working on the whole evolution if you will of RtI for the last 18 months or so. But it has not yet reached a stage where I have become involved to evaluate what's going on because at this stage I don't think that it's evaluable at this point in time. We haven't reached a point of being able to say well we can now start at this baseline and then move on as far as the future is concerned. Did I make myself clear in that? Okay. Any other questions? Yes.

00:34:00

Dr. Hinson: Yeah, yeah we normally have a great response rate from students because they're in of course enclosed environment, and you will fill out the survey. Great response rate there, to give an example on parents. We sent out 60,000 plus surveys to parents this past year, and about 20 percent responded, and that's kind of been the trend for the last few years. That's about 12,000 or so something like that. With respect to, give another example, we, we administered two major surveys a year that encompass major stake holders around January we conduct the division wide survey and that goes to students, teachers, school based administrators, school based non-administrators, and all central office personnel, and parents.

00:35:06

Dr. Hinson: The other survey is what we've just finished now is a customer satisfaction survey. This survey is completed only by school based administrators, and all central office administrators. And what that survey does it seeks to solicit how satisfied individuals are with the services provided by each of the main offices in central office. So the school has a chance to, to chime in, and all central office personnel have a chance to chime in. And for the response rate for that this year we had like 88 percent of central office people responded, and about 70 percent of school based administrators responded to that particular survey. When I go back to, I can't remember the figures for the division wide survey right now, but they were pretty good. We normally have a good representation of what people think and feel across the board.

00:36:05

Dr. Hinson: Yes. Yes exactly, exactly, any other questions? Yes.

00:37:00

Dr. Hinson: Yeah, yeah true. Yeah anybody else? Oh we, okay.

## Response to Intervention

### Cohort training session: Process Evaluation: Is it Working?

00:38:00

Dr. Parker: So I have been with Prince William County for about three years now, and just for the purpose of disclosure I, I as one of the vendors, you know, I was by trade, you know, I'm a psychometrician, so I'm a nerd. And for those of you who are unfamiliar with what that is that's the guy who does the stats and the modeling and all that stuff with the standardized tests, so that's my background. How is that? Good? Okay, so people started to find out what my background was, they started to approach me, and say well can you help us analyze our data. I said, sure, I'm one of those people that actually believes that it takes a village, so I'm always willing to sit down and talk to anybody. So one principal came in and they dropped this binder of data on my desk, and you know, my first reaction was, you know, what the hell am I supposed to do with this?

00:39:04

Dr. Parker: Got all this stuff here right? And it, you know, they're, they're just being honest in how they communicated, and they said well I got to collect data, I got to collect data, I got to collect data. It's like yeah you do, you got to collect data, but what is all this data telling you, and why did you do this, this, this, and this. So I began to have conversations with other people in our division, same thing. Well I got to collect data, got to collect data. Data, data, data, got to use this test, got to use this test, got to use this test. So everybody starts to complain about well I hate the SOLs, or like this test, the one that they hate I hate the SOLs, I never got that one. It's like hating a ruler cause you're not tall enough, you know what I mean? so anyway, so I, I started to apply my, my past with what I was seeing in the schools, and what, where I landed was it really is a question of validity.

00:40:02

Dr. Parker: Right? And validity, everybody in here probably has had some exposure to validity, but from a assessment specialists position it's the most misunderstood thing in education, what is validity? Now if you ask the common person in the class they're going to say when a test measures what it's supposed to measure. Right? We've all probably heard a semblance of that definition before. Okay so let's say you're giving a third grade math test to measure third grade math skills, is that a valid test? It's in the ballpark right? But what if the next day you gave that same test to measure English skills? Doesn't make any sense. Well nothing happened with the test right? It's the decisions that we make based on the information we receive from the test, so first thing to keep in mind validity has to do with the decisions that we make. And those decisions are predicated on what our purpose is from jump.

00:41:03

## Response to Intervention

### Cohort training session: Process Evaluation: Is it Working?

Dr. Parker: Okay, so that's kind of what I want to talk about today is what is validity? And what I'm gonna talk about is there is no other reason to talk about it than to get you thinking okay, and what I'm gonna talk about in no way presupposes that anything you're doing is wrong. Cause a lot of times it's not, but what I do want you to do is begin to think a little bit differently about some of the things you're doing, some of the decisions that you're making, why you're making it, and how you're using information okay. So first of all this is sort of the nerd definition of validity. The degree to which evidence and theory support the interpretation and uses entailed by proposed uses of the test. It's the AERA document, it's kind like the bible for assessment. But this already tells us that validity is not a yes no proposition is it? No it isn't, you don't say that's invalid. I just haven't seen enough decision, enough information to feel good about the decision that I'm going to make that uses this assessment as a source of information.

00:42:06

Dr. Parker: Okay, so how do you evaluate the validity of the decisions that you make that use assessment as a source of information? Well this guy, his name is [Sam Mesk?], he's kind of like the king of the nerds, and he describes six aspects okay. And when he originally described these different aspects of validity he used them specific to educational measurement, right standardized tests. But what I want you to keep in mind is what we're gonna talk about extends beyond standardized tests, it really does include any assessment you make. It's just that how you validate the decision that uses that assessment as a source of information varies according to your purpose and what decision you ultimately are making. Okay, so I'm just gonna show these, these are the six different aspects and what I want you to do is again don't try to write everything down, and don't try to absorb everything.

00:43:05

Dr. Parker: I just want you to start thinking about the types of decisions that you're making about these very special students. And how comfortable you are in making those decisions based on the information you're assessments provide you, okay there is always going to be a content aspect of these decisions you make. Am I asking the right stuff when I make an assessment of a child? But also extends beyond the am I asking the right stuff, it's how well am I asking the right stuff right? You can ask the right stuff the wrong way and get poor information back right? Structural, there is a structural part to assessment, and we have to keep that in mind. Let's say we were giving a benchmark test, or a series of benchmark tests across the course of the year, and we wanted to measure student's progress in say reading, well let's do math that's a little bit easier.

00:44:04

## Response to Intervention

### Cohort training session: Process Evaluation: Is it Working?

Dr. Parker: If we were measuring math skills or assessing math skills we would expect that a child who could subtract could also add, is that a safe assumption? Of course it is, but what if that doesn't manifest in our test scores? Something is wrong right? We don't really want to know A through Z on a test, we think when we talk about it that's what we say we want to know this stuff. We really want to know how well a child knows the alphabet right? And we use A through Z as a manifest indicator of how well the child knows the alphabet. So what we want to know is already in the kid's head, it's in his squish. What's in the child's squish should manifest in test responses or assessment responses, and if they don't there is a disconnect. Either our theory is terrible, or the test doesn't reflect our theory to the degree it should.

00:45:01

Dr. Parker: Those are things we can check, I'll talk a little bit about how we can do that in a minute. Substantive, we think that when we assess a child we know what's going on inside the child's head right? But how often do we check that? And if you're using assessments that publishers have provided for you have they shown you evidence that allows you to conclude yes what I think is going on in the child's head is, is actually what's going on in the child's head. The publisher if he's worth his salt should provide you with some information on that. Right? I mean if you have been a classroom teacher you'll ask a question, I mean the kid could be thinking about I don't know I'm gonna make some spit balls. Right, and something gets blurted out and the kid misses the question or whatever. What you think happened, or what you intended to happen didn't actually happen. There is ways to check for that, and if we can confirm that we can feel better about the information the assessment has provided for us.

00:46:03

Dr. Parker: Generalizability, this is a big one, right this is reliability. First of all tests are reliable, data are reliable right? Will data replicate over time, that's one way to look at how generalizable are my results. If I give this assessment on this day will I expect to see the same rank ordering of students the next day? And then the next day without there being some type of effect, does everybody track that? But it also extends beyond the time in which you give the test, it also includes different types of populations, and this is where it's important for you. You're dealing with very specialized populations. Can we use this assessment on those different populations and still feel, am I zapping people with this red? Sorry I'll zap myself. So there is an external aspect and I'll talk a little bit more about that in a second.

00:47:04

Dr. Parker: And obviously there is a, maybe not obviously the most neglected part of assessment is the consequential aspect. What are the consequences to giving this assessment? This is a touch subject for me, so anyway the content aspect. The

## Response to Intervention

### Cohort training session: Process Evaluation: Is it Working?

question you want to ask before you select an assessment or before you create an assessment is the assessment content relevant, or irrelevant and representative of the subject area I'm interested in. How can you check this? Well if you're creating a test you can go through these steps, or if you're using a test or an assessment from a publisher again if he or she is worth their sand they'll provide you a technical manual that provides information about this. If they don't you have to question why. Again I was a test publisher, so I'm not, if you're worth your sand what do you have to hide, this information should be available to you.

00:48:01

Dr. Parker: Purpose, you should all start with a purpose. Going back to my example earlier, when that principal laid this enormous, by the way it wasn't [Mr. Jacks ?], [Mr. Jacks ?] is excellent. He really is, we've, we had a nice meeting recently. But when this principal, what else am I gonna say. [LAUGHTER] When I, he fills out our customer satisfaction survey as far as he knows it was a great meeting, no it really was yeah it was great. Anyway when I talked to this principal they laid all this data out for me, and I said well they said man this is awesome this is from this test. This is from this test, this is from this test. And I asked him well why did you do all this stuff? And what he responded back with, it sounded like the definition the publisher, the test publisher had given him. Well this test measures bah, bah, bah, bah, bah, I'm like man I don't want to know that. I want to know why you used this test, you're the theory, you have to have a purpose.

00:49:00

Dr. Parker: Don't the tail wag the dog, start out with a purpose. Go out and find the assessment that meets your purpose with the information you need. Don't be force fed something, if you do you're missing the point. You know your children the test publisher does not know your children. Know your children, have a purpose, seek out the assessment that maps to that. There should be specifications you guys know the stuff right? Or you teachers know the stuff, they know that construct. If you're using something that was created by somebody else there should be specifications in there that show you whether or not they're test, or their assessment matches the construct as you envision it. If it doesn't you got to move on, you got to create something yourself. The process through which the assessment was created. If you're using a scoring rubric there should be information about that process.

00:50:01

Dr. Parker: If it's a test, a multiple choice item test for instance, they should talk about they pre-tested the questions. Remember how well do we ask things? Again this applies to all assessments. There should be some expert review by people, content experts such as yourself, or teachers in your classrooms. There should be some

## Response to Intervention

### Cohort training session: Process Evaluation: Is it Working?

sensitivity reviews, again back to that generalizability do we predict that this question will function differently for a student with a disability versus a child without a disability. There should be some evidence in that technical manual that talks about that, if it isn't you got to wonder why. Maybe the test publisher didn't, didn't know any better. Well then, there should be some information about piloting and field testing of the assessment whatever it is. And they should also talk about the technical quality of the questions, are the questions too hard for my population?

00:51:01

Dr. Parker: Is the cognitive level required by this assessment beyond the cognitive level of my population that I'm going, again these are things you have to consider. Otherwise you can't have confidence in the decisions that you ultimately make that use the assessment as a source of information. So the structural aspect, again this goes back to our add, subtract example. The test publisher or you should be doing some exploration into the dimensionality of the assessment. For instance if we're giving a benchmark assessment we usually give a pass/fail, or a single numerical score right? Right? Yeah, okay so a single score we'd better be measuring a single thing. What if we're measuring two things? You no longer are able to interpret that score. What if the language demands of the test exceed the language ability of an ELL population, and you're still testing that ELL population with that benchmark test?

00:52:01

Dr. Parker: It's no good. There is now multiple sources of variability in your responses. You can't interpret that test score correctly, does everybody know what I mean. Yeah it jacks everything up, you know, you can't use it. If you're using an assessment from a publisher there should be information about correlations between test questions. Again pronouns, student performance on questions about pronouns would logically be correlated with student performance on synonyms I guess. I stink at language arts, but maybe math is better. Adding and subtracting, you would expect those to be correlated, and if they're not something has happened right? The test isn't functioning as we intended. There should also be some information around dimensionality analysis. Incidentally I know some of these are really abstract terms, I'll give you a card. I really do believe it takes a village call me up any time we can talk about it.

00:53:03

Dr. Parker: substantive aspect. Is the child engaging in the cognitive process we believe while he or she participates in the assessment? How do we know? Well there is ways to check, and this is probably the easiest one. If we're engaged with our students we should know this anyway okay, watch them when they're completing the assessment watch them. Look at eye movement okay, take two or three kids and ask them to complete the assessment out loud, ask them to verbally process as they go

## Response to Intervention

### Cohort training session: Process Evaluation: Is it Working?

through the assessment. You'd be surprised at some of the stuff that flies out of their mouth. They're not engaged in the process that we necessarily believe, and if they are let's confirm that. Again we're building an evidence base so that we feel good about the conclusions we ultimately make.

00:54:00

Dr. Parker: Look at scratch paper. I'm not saying walk out of the room with SOL scratch paper, that's, no. But in your small scale assessments, if they're allowed, I know a lot of us do practice SOL tests right? Have them use their scratch paper replicating the assessment environment, look at it. You know, if they're, you know, drawing checkerboards and stuff they're probably not engaged in the process we envisioned. If you're creating tests with teachers in the schools, an interesting thing to do is have the teachers predict how hard the questions are before they give the test, and the match that to how hard the questions actually were. If they match pretty closely well then what you've done is you've shown that the teachers who helped create the test were really in tune with not only their students but with the content.

00:55:00

Dr. Parker: Right? We again, we feel good about the results and the decisions we make that use this assessment as a source of information. The generalizability, these are fancy words, basically does the assessment hold over different contexts? We're already talked about the reliability one. For those of you who aren't statisticians, I'm guessing zippy in the room and that's cool. Reliability is a number that varies between zero and one, the closer you get to one the better. The SOLs are in that .8 to .93 range, that's pretty good. I feel pretty good about that. Look for these things in the technical manual. If you're using another assessment look for reliability indices, at minimum they should report that if they don't you got to question why. This confirm item in variance, export diff, diff is called differential item functioning, that's a super fancy word of saying a question acts differently for an ELL versus a non-ELL.

00:56:11

Dr. Parker: Or a question acts differently for a student with a disability versus a child without a disability. If you are going to use these assessments on different populations you better gosh darn be sure that the question doesn't function differently for them. Otherwise your results are jacked right, you can't be as confident in the results if you're using for ELLs versus non-ELLs etc. Okay, and a publisher worth his salt should have some information in there about how they explored this. A sensitivity review is sort of a way to predict whether a question will function differently, and I'm not just talking about pass rates right. I mean there are reasons why some populations don't do as well as others.

## Response to Intervention

### Cohort training session: Process Evaluation: Is it Working?

00:57:00

Dr. Parker: Things that don't have anything to do with a test, but if the test is causing this differential functioning to occur you got to address it. The external, I won't talk too much about this other than to say you would expect the assessment results that you receive to match constructs, similar constructs that exist. One simple way I heard a teacher do it, they created a classroom test, they matched it to their grade book. If the results are really discrepant, either the grading system is whacked, which is probably the case. Or the assessment is whacked, but you could look at okay did the kids on this test do about the same goodness as the kids on the SOL? Right, you can sort of validate the different assessments by matching them up to each other, that's what this is saying. Finally consequences, you need to think about the consequences before you give these assessments.

00:58:04

Dr. Parker: Okay, there is a, again that's why I went into math, there is a sociological aspect to testing. There is a psychological aspect to testing, what's the impact going to have on this child? What's the impact going to have on his instructional plan moving forward? Do I haphazardly use this assessment and blurt out the decision? You got to think about all these types of things. What are some different ways you can check this? Well for one thing you can make sure if you are giving say a benchmark test your cut scores are in the right place. Okay, especially if you're comparing, let's say you were doing a math benchmark over time. If you're comparing pass rates over time, how many people here have taken benchmark tests for a child and compared scores across time?

00:59:03

Dr. Parker: Come on man, yeah you have of course. How many people here have considered how difficult each form of the test was before doing that? It's different information isn't it, right? A pass rate on a hard test is not the same as a pass rate on an easy test. So how can you compare those two pass rates? Right? One of the things you can do is make your cut scores are set in the right place, some very simple methods if you are doing benchmark testing in your division that you can incorporate to have more confidence, and it won't be perfect. But more confidence in where you set your passing scores. Let me tell you that saying 70 percent correct on the test is no good. Again that doesn't account for how easy the test is versus how easy the test is. Okay, okay.

01:00:00

Dr. Parker: So some final comments on validity, and again assessment in general. Evaluating the validity of decisions about kids extends beyond any single assessment,

## Response to Intervention

### Cohort training session: Process Evaluation: Is it Working?

this is the multiple measures concept. But multiple measures does not just include multiple tests right? Reliability has to do with replicating results, so there is different ways you can do that. Different modes of assessment, if people are giving standardized tests or multiple choice every Friday and that's it, come on man. Right? Frequency of assessment, the appropriateness of assessment different reasons why we do different things. But again it all starts with your purpose. You have to link what you're doing back to your purpose. Validation is an iterative process, it's a continuous process. If you are using tests, and I mean you know, we are, for a lot of people we're talking about a paradigm shift.

01:01:02

Dr. Parker: You know, if I'm doing something it must be the right thing cause I'm doing it. People are very comfortable with the stuff they're doing right? But what about the test or the assessment tool that we've been using for 20 years? You know what? You're right things haven't changed over the last 20 years, keep doing what you're doing. There is a reason why a lot of these big tests that come out with different norming books, but they should also come out with new technical manuals. To the state's credit they do a lot of these things with the SOLs, and they do publish technical manuals. Okay, and again different sources of validity evidence should be, yes sir.

01:02:00

Dr. Parker: I would recommend using, I would, the process, there are items centered processes, and what you do is you decompose the test you create and you look at each individual question in terms of how difficult. There is something called an Angoff it's a very simple process you look at one question, you predict the percentage of borderline students that would get that right, and then you look at the next question, borderline how many, what percentage of borderline kids?

01:03:03

Dr. Parker: And then what you end up is you add those up and you get your cut score, and then you readjust it according to some empirical data. Seriously give me a call I'll give you my card, I'll give you a reference. It really is an easy process. See what's funny is all the nerds get together and they use the fancy language to turn you off, and make you think that they have insight to something you don't. The truth of the matter is that you, you know the kids right? If anybody here has ever participated in standard setting it's not the nerds setting the standards, it's them bringing in a team of qualified teachers because they're the ones who's brains have to be picked. Cause they know the kids and they know the content. So there is a simple process you could go through. That's it, that's what the state uses. I wouldn't, don't get freaked out it's pretty easy.

01:04:00

## Response to Intervention

### Cohort training session: Process Evaluation: Is it Working?

Dr. Parker: So any questions? Yes sir. Yeah I think so, I mean we have conversations. I mean we're a big division man, and we've got, you know, we're 88 schools. So we've got, I mean we run the gamut in terms of ability and what people think they know too, you know what I mean? And like I mean [Mr. Jacks ?] and I, I was being serious had a very nice conversation about three or four months ago where that's exactly what we did.

01:05:01

Dr. Parker: We just bounced stuff off of each other, and he kind of settled. He knows the kids, you know, he knows the kids. And I, I'd like to think that we have those, those types of conversations with all of our principals. You know, the thing that I often, what I find myself repeating with people most often though is well what do you want to know? What's your purpose? And that's what makes, people are like what? You know, cause a lot of times it is well what did the publisher tell me. And a lot of times they're very robotic responses research bases blah, blah, blah. I think usually through these conversations as you, you, once we define a purpose everything else kind of falls in place, you know, it, it more of less naturally flows.

01:06:08

Dr. Parker: My experience thus far and Ken may have a different experience, but my experience thus far is really defining first and foremost what's the purpose. What do you want to know? Right man people's heads are spinning, they got SOLs, they got benchmark tests, they got I mean, it's tough man. Well everything is important cause it's what they do, you know what I mean?

01:07:00

Dr. Parker: Yeah parent wants their kid to read. And I bet you most often if we're doing our jobs and we teach the kids to read the SOL takes care of itself.

01:08:00

Dr. Parker: Well that's it, that's the whole theory of multiple measures. It's some point, you know, I asked my wife what she wants to do on Saturday on Monday, she tells me the same thing on Tuesday. She tells me the same thing on Wednesday, Thursday she tells me to shut up and leave her alone. You know what I mean? you reach that point of saturation, and but one time, to be honest my wife changes her mind a lot. So I'm going to ask her again on Tuesday. You know, we're talking about the reliability of the information we get. I'm serious. It's, you know, it's a, it's a moderate example, but if she tells me on Tuesday and Wednesday I start to buy it, and on Thursday enough is enough. I'm not wasting my time and ticking her off, so I'll stop it. What you're describing is, is the case, but again when we go back to our validity definition it's at what

## Response to Intervention Cohort training session: Process Evaluation: Is it Working?

point do you feel confident, you know, the point at which I feel confident may be different from the point at which you feel confident.

01:09:04

Dr. Parker: And those are the conversations that I have with our principals is that those are individual decisions for individual students. What I really appreciate about our conversation is how you were treating the child as an individual, and how you were creating a plan for that individual child. That's, I really, really liked that, so that's kind of what I'm talking about, decisions about an individual child, valid for that individual child, etc.

01:10:00

Dr. Cave: [Tape cuts] --page and then the three colored ones are the emerging, the developing and the sustaining, so we're going to be passing those out in case you didn't pick them up. [Tape cuts]

Dr. Hinson: For everybody's information we originally planned to give you some time to do an activity, I have passed that activity out to each table.

01:11:01

Dr. Hinson: But my admin assistants didn't follow instructions totally the way I wanted. You should have had separate sheets, and you should only have had the blank sheet first, and the back sheet second. But that's quite all right, the whole idea of this particular exercise was to allow you to think about your own divisional situations with respect to an evaluation plan. Can you hear me, and as I said before we are relatively large school district, so we have the resources available to do this sort of thing. Some of you may not have those kinds of resources, so what I'm, I'm only asking you to do is to look at the particular benchmark here listed for evaluation, the evaluation plan has multiple components and includes assessing both short and long term impact of the TSS implementation. Again what I said before in my presentation is, is how do you know you're meeting that benchmark? And what kinds of questions do you need to pose in your own mind with respect to your given community or division that will allow you to answer that question.

01:12:06

Dr. Hinson: Now on the back you can already see that I just brainstormed any number of questions that could be posed, and you may look at that and say well hey these are some of the things that I would think about, but there are others that I would think about. But that was the whole idea of this planned activity, and right now we're about 15 minutes of noon, and I would say Regina just let them review it as is. And cause we're

## Response to Intervention Cohort training session: Process Evaluation: Is it Working?

bout ten minutes before lunchtime right now. Yeah, yeah, yeah. Yeah. Yes, yes again the activity sheet that I have passed out to everyone, the idea behind this activity is this, you saw the presentation I made where I had a very large question at hand, and I went through a series of breaking it down into smaller questions that could be posed in order to answer the larger question.

01:13:03

Dr. Hinson: In this particular case of all of the benchmarks listed, one that I'm talking about here is the evaluation plan has multiple components and includes assessing both short and long term impact of the TSS implementation. So I'm saying that within your own divisions or sea of operation think in terms of what you would come up with as far as questions go to say yes we are meeting this benchmark? Kind of the same kind of thought process I had before, and as I said before my admin assistant made the mistake of clipping these two pages together. You were only to get the blank page first, and I was going to give you the second page later. But again I have listed any number of questions that could be posed, should be posed to answer that larger question, and they maybe some that would fit what you were doing in your own environment, you may have others.

01:14:00

Dr. Hinson: And the whole idea right now is to basically look at what I have, and what you would think, and how it would fit in your own division in trying to get to the point of are we actually meeting what we think we need to do with respect to having a valid evaluation plan for the TSS? Make sense? Okay. [Tape cuts] --screen you should have it in your handouts, and literally if you have any more questions or concerns that you want to pose to either me or to Paul feel free to give us a call or send us an email, no problem. Thank you.

01:15:00

Dr. Cave: It's lunchtime.